

Covariate adjustment in randomization-based causal inference for 2^K factorial designs

Jiannan Lu*

Analysis and Experimentation, Microsoft Corporation

Abstract

We develop finite-population asymptotic theory for covariate adjustment in randomization-based causal inference for 2^K factorial designs. In particular, we confirm that both the unadjusted and the covariate-adjusted estimators of the factorial effects are asymptotically unbiased and normal, and the latter is more precise than the former.

Keywords: Potential outcome; Variance reduction; Finite-population asymptotics

1. INTRODUCTION

Randomization is often considered the gold standard for causal inference (Rubin 2008). A well-established methodology to conduct causal inference is the potential outcomes framework (Neyman 1923; Rubin 1974), which defines the causal effect of a binary treatment factor as the comparison between the potential outcomes under treatment and control. In the presence of multiple binary treatment factors, we can evaluate them simultaneously under the 2^K factorial design framework (Fisher 1935; Yates 1937). Several researchers (e.g., Kempthorne 1952, 1955; Wilk and Kempthorne 1956; Bailey 1981, 1991; Dasgupta et al. 2015) advocated conducting randomization-based causal inference for 2^K factorial designs, which has several advantages over the widely-used regression-based inference. For example, randomization-based inference is applicable to the finite-population

*Address for correspondence: Jiannan Lu, Microsoft Corporation, One Microsoft Way, Redmond, Washington 98052, USA. Email: jiannl@microsoft.com

setting, and therefore may be more reasonable in practice (e.g., Miller 2006; Lu et al. 2015). For more discussion on the comparison and reconciliation of randomization-based and regression-based inferences for 2^K factorial designs, see Lu (2016).

In randomization-based causal inference, covariate adjustment (Cochran 1977) is a variance reduction technique widely used by researchers (e.g., Deng et al. 2013; Miratrix et al. 2013). In an illuminating paper, Lin (2013) demonstrated the advantages of performing covariate adjustment for randomized treatment-control studies (i.e., 2^1 factorial designs). However, to our best knowledge, for 2^K factorial designs which are of great importance from both theoretical and practical perspectives, similar discussions appear to be absent; it is unclear whether covariate adjustment is beneficial for 2^K factorial designs, and if so, how to quantify said benefit. In this paper we answer this question, by extending the discussions in Lin (2013) and illustrating the advantages of performing covariate adjustment in 2^K factorial designs. To be specific, we derive the closed-form expressions for the asymptotic precisions of the unadjusted and covariate-adjusted estimators, and thus accurately measure the precision gained by covariate adjustment.

The paper proceeds as follows. Section 2 reviews randomization-based inference for 2^K factorial designs. Section 3 introduces the covariate-adjusted estimator for 2^K factorial designs. Section 4 derives the asymptotic precisions of the unadjusted and covariate-adjusted estimators. Section 5 concludes and discusses possible future directions.

2. RANDOMIZATION INFERENCE FOR 2^K FACTORIAL DESIGNS

In this section, we review the randomization-based inference framework for 2^K factorial designs (Dasgupta et al. 2015; Lu 2016). For consistency we adopt the notations in Lu (2016).

2.1. 2^K factorial designs

2^K factorial designs consist of K distinct treatment factors, each of which has two levels coded as -1 and 1. To simplify future notations we let $J = 2^K$. To define 2^K factorial designs, we rely on a $J \times J$ orthogonal matrix $\mathbf{H} = (\mathbf{h}_0, \dots, \mathbf{h}_{J-1})$, which is often referred to as the model matrix (Wu and Hamada 2009). We construct the model matrix in the following recursive way (Espinosa et al. 2016; Lu 2016):

1. Let $\mathbf{h}_0 = \mathbf{1}_J$;
2. For $k = 1, \dots, K$, construct \mathbf{h}_k by letting its first 2^{K-k} entries be -1, the next 2^{K-k} entries be 1, and repeating 2^{k-1} times;
3. If $K \geq 2$, order all subsets of $\{1, \dots, K\}$ with at least two elements, first by cardinality and then lexicography. For $k = 1, \dots, J - 1 - K$, let σ_k be the k th subset and $\mathbf{h}_{K+k} = \prod_{l \in \sigma_k} \mathbf{h}_l$, where “ \prod ” stands for entry-wise product.

The j th row of the sub-matrix $\tilde{\mathbf{H}} = (\mathbf{h}_1, \dots, \mathbf{h}_K)$ is the j th treatment combination \mathbf{z}_j . To further illustrate the construction of the model matrix, we adopt the example in Lu (2016).

Example 1. Let $K = 2$. By following the above recursive procedure, we obtain $\mathbf{h}_0 = \mathbf{1}$, $\mathbf{h}_1 = (-1, -1, 1, 1)'$, $\mathbf{h}_2 = (-1, 1, -1, 1)'$, and $\mathbf{h}_3 = (1, -1, -1, 1)'$. Consequently, for 2^2 factorial designs the model matrix is:

$$\mathbf{H} = \begin{matrix} & \mathbf{h}_0 & \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \\ \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \end{matrix}.$$

The four treatment combinations are $\mathbf{z}_1 = (-1, -1)$, $\mathbf{z}_2 = (-1, 1)$, $\mathbf{z}_3 = (1, -1)$ and $\mathbf{z}_4 = (1, 1)$.

2.2. Randomization-based Inference

We allow $N \geq 2J$ experimental units in the design. To describe the randomization-based inference framework, we follow a three-step procedure.

First, under the Stable Unit Treatment Value Assumption (Rubin 1980) that for $j = 1, \dots, J$ there is only one version of the treatment combination \mathbf{z}_j , and no interference among the experimental units, let $Y_i(\mathbf{z}_j)$ be the potential outcome of unit i under treatment combination \mathbf{z}_j , and $\bar{Y}(\mathbf{z}_j) = N^{-1} \sum_{i=1}^N Y_i(\mathbf{z}_j)$ be the average potential outcome across all the experimental units. Let $\mathbf{Y}_i = \{Y_i(\mathbf{z}_1), \dots, Y_i(\mathbf{z}_J)\}'$ and $\bar{\mathbf{Y}} = \{\bar{Y}(\mathbf{z}_1), \dots, \bar{Y}(\mathbf{z}_J)\}'$.

Next, we randomly assign $n_j \geq 2$ units to treatment combination \mathbf{z}_j . Let

$$W_i(\mathbf{z}_j) = \begin{cases} 1, & \text{if unit } i \text{ is assigned treatment } \mathbf{z}_j, \\ 0, & \text{otherwise,} \end{cases}$$

and let $Y_i^{\text{obs}} = \sum_{j=1}^J W_i(\mathbf{z}_j) Y_i(\mathbf{z}_j)$ be the observed outcome for unit i , and therefore the average observed outcome across all experimental units that are assigned to treatment combination \mathbf{z}_j is $\bar{Y}^{\text{obs}}(\mathbf{z}_j) = n_j^{-1} \sum_{i=1}^N W_i(\mathbf{z}_j) Y_i(\mathbf{z}_j)$. Furthermore, we let $\bar{\mathbf{Y}}^{\text{obs}} = \{\bar{Y}^{\text{obs}}(\mathbf{z}_1), \dots, \bar{Y}^{\text{obs}}(\mathbf{z}_J)\}'$.

Finally, we define the factorial effects as

$$\tau(l) = \frac{1}{2^{K-1}} \mathbf{h}_l' \bar{\mathbf{Y}} \quad (l = 1, \dots, J-1),$$

and their randomization-based estimators as

$$\hat{\tau}_{\text{rb}}(l) = \frac{1}{2^{K-1}} \mathbf{h}_l' \bar{\mathbf{Y}}^{\text{obs}} \quad (l = 1, \dots, J-1). \quad (1)$$

Its randomness is solely from the treatment assignment $W_i(\mathbf{z}_j)$'s.

3. COVARIATE ADJUSTMENT IN 2^K FACTORIAL DESIGNS

The idea behind the randomization-based estimator is estimating the average potential outcome $\bar{Y}(\mathbf{z}_j)$ by its corresponding average observed outcome $\bar{Y}^{\text{obs}}(\mathbf{z}_j)$. However, as shown in Cochran (1977) and later mentioned in Lin (2013), utilizing the pre-treatment covariates can potentially improve the precision of $\bar{Y}^{\text{obs}}(\mathbf{z}_j)$, and consequently that of the randomization-based estimator. With this classic wisdom, we define the covariate-adjusted estimator for 2^K factorial designs. In this paper, we consider the method of covariate adjustment where separate slope coefficients are estimated for each average potential outcome $\bar{Y}(\mathbf{z}_j)$, unlike the traditional ANCOVA method in which there is only one pooled slope coefficient. The rationale behind this is from the existing literature on covariate adjustment in randomized treatment-control studies – as shown in Freedman (2008) and Lin (2013), the traditional ANCOVA can potentially help or hurt asymptotic precision, however the “separate slope” method guarantees asymptotic precision improvement.

Let $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})'$ be the pre-treatment covariates of the unit i , and $\bar{X}_k = N^{-1} \sum_{i=1}^N X_{ik}$ and $\bar{X}_k^{\text{obs}}(\mathbf{z}_j) = n_j^{-1} \sum_{i=1}^N W_i(\mathbf{z}_j) X_{ik}$ be the average of the k th covariate of all units and those assigned to treatment \mathbf{z}_j . Let $\bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_p)'$, and $\bar{\mathbf{X}}^{\text{obs}}(\mathbf{z}_j) = \{\bar{X}_1^{\text{obs}}(\mathbf{z}_j), \dots, \bar{X}_p^{\text{obs}}(\mathbf{z}_j)\}'$. Consider the following type of estimators for $\bar{Y}(\mathbf{z}_j)$:

$$\bar{Y}^{\text{obs}}(\mathbf{z}_j) + \{\bar{\mathbf{X}} - \bar{\mathbf{X}}^{\text{obs}}(\mathbf{z}_j)\}' \boldsymbol{\beta}_j,$$

where $\boldsymbol{\beta}_j$ is a constant vector to be determined. As shown in Cochran (1977), the value of $\boldsymbol{\beta}_j$ that minimizes the variance of the above is

$$\boldsymbol{\beta}_j = \left\{ \frac{1}{N} \sum_{i=1}^N (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})' \right\}^{-1} \left[\frac{1}{N} \sum_{i=1}^N (\mathbf{X}_i - \bar{\mathbf{X}}) \{Y_i(\mathbf{z}_j) - \bar{Y}(\mathbf{z}_j)\} \right], \quad (2)$$

which we assume to be well-defined, i.e., the “design matrix” $N^{-1} \sum_{i=1}^N (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})'$ is invertible. We estimate (2) by the plug-in method:

$$\hat{\boldsymbol{\beta}}_j = \left\{ \frac{1}{N} \sum_{i=1}^N (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})' \right\}^{-1} \left[\frac{1}{n_j} \sum_{i=1}^N W_i(\mathbf{z}_j) (\mathbf{X}_i - \bar{\mathbf{X}}) \{Y_i(\mathbf{z}_j) - \bar{Y}^{\text{obs}}(\mathbf{z}_j)\} \right], \quad (3)$$

and let

$$\bar{Y}^{\text{ca}}(\mathbf{z}_j) = \bar{Y}^{\text{obs}}(\mathbf{z}_j) + \{\bar{\mathbf{X}} - \bar{\mathbf{X}}^{\text{obs}}(\mathbf{z}_j)\}' \hat{\boldsymbol{\beta}}_j \quad (j = 1, \dots, J). \quad (4)$$

Consequently, we define the covariate-adjusted estimator as

$$\hat{\tau}_{\text{ca}}(l) = \frac{1}{2^{K-1}} \mathbf{h}_l' \bar{\mathbf{Y}}^{\text{ca}} \quad (l = 1, \dots, J-1), \quad (5)$$

where $\bar{\mathbf{Y}}^{\text{ca}} = \{\bar{Y}^{\text{ca}}(\mathbf{z}_1), \dots, \bar{Y}^{\text{ca}}(\mathbf{z}_J)\}'$.

4. FINITE-POPULATION ASYMPTOTIC ANALYSIS

4.1. Notations and Assumptions

Consider a hypothetical sequence of finite populations with increasing sample sizes. Technically, all the finite-population quantities should have superscripts that index the sequence of populations, for example $\bar{\mathbf{X}}^{(N)} = N^{-1} \sum_{i=1}^N \mathbf{X}_i$. For convenience we drop all superscripts. We make the following

assumptions to conduct the finite-population asymptotic analysis, and the first assumption merely serves the purpose of simplifying notations.

Assumption 1. For all N , the potential outcomes and covariates are centered at zeros, i.e., $\bar{\mathbf{X}} = 0$, and $\bar{\mathbf{Y}} = 0$.

Assumption 2. When $N \rightarrow \infty$, the proportions of units assigned to all treatment combinations converge to positive constants, i.e.,

$$\hat{p}_j = \frac{n_j}{N} \rightarrow p_j \quad (j = 1, \dots, J),$$

where $p_j > 0$ for all j and $\sum_{j=1}^J p_j = 1$.

Assumption 3. When $N \rightarrow \infty$, all the second moments of the potential outcomes converge to constants, i.e.,

$$\frac{1}{N} \sum_{i=1}^N \mathbf{Y}_i \mathbf{Y}_i' \rightarrow \boldsymbol{\Sigma} = (\sigma_{jj'})_{1 \leq j, j' \leq J}$$

where $\sigma_{jj} > 0$ for all j . All the second moments of the covariates converge to constants, i.e.,

$$\frac{1}{N} \sum_{i=1}^N \mathbf{X}_i \mathbf{X}_i' \rightarrow \boldsymbol{\Omega} = (\omega_{ll'})_{1 \leq l, l' \leq p},$$

where $\boldsymbol{\Omega}$ is an invertible matrix. All the mixed second moments of the potential outcomes and the covariates converge to constants, i.e.,

$$\frac{1}{N} \sum_{i=1}^N \mathbf{X}_i Y_i(\mathbf{z}_j) \rightarrow \boldsymbol{\lambda}_j \quad (j = 1, \dots, J).$$

Assumption 4. For all N , the fourth moments of the potential outcomes and the covariates are uniformly bounded from above by a positive constant, i.e.,

$$\frac{1}{N} \sum_{i=1}^N Y_i^4(\mathbf{z}_j) \leq L \quad (j = 1, \dots, J); \quad \frac{1}{N} \sum_{i=1}^N X_{ik}^4 \leq L \quad (k = 1, \dots, p).$$

We introduce several useful notations before moving forward. Let

$$\boldsymbol{\zeta}_j = \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}_j \quad (j = 1, \dots, J); \quad R_i(\mathbf{z}_j) = Y_i(\mathbf{z}_j) - \mathbf{X}_i' \boldsymbol{\zeta}_j \quad (i = 1, \dots, N),$$

and $\mathbf{R}_i = \{R_i(z_1), \dots, R_i(z_J)\}'$. Consequently,

$$\frac{1}{N} \sum_{i=1}^N \mathbf{R}_i \mathbf{R}_i' \rightarrow \tilde{\Sigma} = (\tilde{\sigma}_{jj'})_{1 \leq j, j' \leq J}, \quad (6)$$

where $\tilde{\sigma}_{jj'} = \sigma_{jj'} - \boldsymbol{\lambda}_j' \boldsymbol{\Omega}^{-1} \boldsymbol{\lambda}_{j'}$.

4.2. Useful Lemmas

For finite-population asymptotic analysis of the randomization-based estimator and the covariate-adjusted estimator, we rely on the following lemmas, which are also of independent interests. The first lemma is the Combinatorial Central Limit Theorem from Hoeffding (1951).

Lemma 1. For fixed $N \in \mathbb{Z}^+$, and N^2 constants a_{jk} ($j, k = 1, \dots, N$), let

$$b_{jk} = a_{jk} - \frac{1}{N} \sum_{j'=1}^N a_{j'k} - \frac{1}{N} \sum_{k'=1}^N a_{jk'} + \frac{1}{N^2} \sum_{j'=1}^N \sum_{k'=1}^N a_{j'k'}. \quad (7)$$

Furthermore, let (ν_1, \dots, ν_N) be a random permutation of $(1, \dots, N)$ and $S = \sum_{j=1}^N a_{j, \nu_j}$. If

$$\lim_{N \rightarrow \infty} \max_{1 \leq j, k \leq N} b_{jk}^2 / \left(\frac{1}{N} \sum_{j=1}^N \sum_{k=1}^N b_{jk}^2 \right) = 0, \quad (8)$$

then when $N \rightarrow \infty$,

$$\frac{S - E(S)}{\{\text{Var}(S)\}^{1/2}} \xrightarrow{\mathbb{D}} N(0, 1).$$

The second lemma is essentially the “ 2^K factorial design version” of the multivariate finite-population Central Limit Theorem in Freedman (2008). However, we provide a rigorous proof in this paper, where Freedman (2008) did not.

Lemma 2. When $N \rightarrow \infty$,

$$N^{1/2} \bar{\mathbf{Y}}^{\text{obs}} \xrightarrow{\mathbb{D}} N(0, \boldsymbol{\Sigma}^{\text{obs}}),$$

where

$$\Sigma^{\text{obs}} = \begin{bmatrix} \frac{1-p_1}{p_1}\sigma_{11} & -\sigma_{12} & \dots & -\sigma_{1J} \\ -\sigma_{21} & \frac{1-p_2}{p_2}\sigma_{22} & \dots & -\sigma_{2J} \\ \vdots & \vdots & \ddots & \vdots \\ -\sigma_{J1} & \dots & \dots & \frac{1-p_J}{p_J}\sigma_{JJ} \end{bmatrix}.$$

Proof. By Cramer-Wold theorem, we only need to prove that

$$N^{1/2}\mathbf{t}'\bar{\mathbf{Y}}^{\text{obs}} \xrightarrow{\mathbb{D}} N(0, \mathbf{t}'\Sigma^{\text{obs}}\mathbf{t}) \quad (9)$$

for all $\mathbf{t} = (t_1, \dots, t_J)' \in \mathbb{R}^J$. If $\mathbf{t} = \mathbf{0}_J$, (9) holds trivially. Otherwise $\mathbf{t} \in \mathbb{R}^J \setminus \{\mathbf{0}_J\}$:

First, by simple probability argument (e.g., Lu 2016, Lemma 1),

$$E(\bar{\mathbf{Y}}^{\text{obs}}) = \mathbf{0}_J; \quad \text{Var}\{\bar{Y}^{\text{obs}}(\mathbf{z}_j)\} = \frac{1-\hat{p}_j}{\hat{p}_j} \frac{1}{N(N-1)} \sum_{i=1}^N Y_i^2(\mathbf{z}_j) \quad (j = 1, \dots, J), \quad (10)$$

and

$$\text{Cov}\{\bar{Y}^{\text{obs}}(\mathbf{z}_j), \bar{Y}^{\text{obs}}(\mathbf{z}_{j'})\} = -\frac{1}{N(N-1)} \sum_{i=1}^N Y_i(\mathbf{z}_j)Y_i(\mathbf{z}_{j'}) \quad (j \neq j').$$

Therefore when $N \rightarrow \infty$,

$$E(N^{1/2}\mathbf{t}'\bar{\mathbf{Y}}^{\text{obs}}) = 0, \quad \text{Var}(N^{1/2}\mathbf{t}'\bar{\mathbf{Y}}^{\text{obs}}) \rightarrow \mathbf{t}'\Sigma^{\text{obs}}\mathbf{t}, \quad (11)$$

Next, we prove that

$$\lim_{N \rightarrow \infty} \frac{\max_{1 \leq i \leq N} Y_i^2(\mathbf{z}_j)}{\sum_{i=1}^N Y_i^2(\mathbf{z}_j)} = 0 \quad (j = 1, \dots, J). \quad (12)$$

Let

$$\eta_i = Y_i^2(\mathbf{z}_j) / \left\{ \sum_{i=1}^N Y_i^2(\mathbf{z}_j) \right\} \quad (i = 1, \dots, N),$$

and obviously $\sum_{i=1}^N \eta_i = 1$. Furthermore, let $\eta = \max_{1 \leq i \leq N} \eta_i$, and consequently

$$\eta \leq \left(\sum_{i=1}^N \eta_i^2 \right)^{1/2} = N^{-1/2} \left\{ \frac{1}{N} \sum_{i=1}^N Y_i^4(\mathbf{z}_j) \right\}^{1/2} / \left\{ \frac{1}{N} \sum_{i=1}^N Y_i^2(\mathbf{z}_j) \right\}.$$

Therefore by Assumptions 3 and 4

$$\limsup_{N \rightarrow \infty} N^{1/2} \eta \leq L^{1/2} / \sigma_{jj},$$

which implies (12).

Then, we adopt the notations in Lemma 1 and let

$$a_{gi} = N^{1/2} \begin{cases} t_1 Y_i(z_1) / n_1, & \text{for } 1 \leq g \leq n_1, \\ \vdots & \\ t_J Y_i(z_J) / n_J, & \text{for } \sum_{j=1}^{J-1} n_j + 1 \leq g \leq N. \end{cases} \quad (i = 1, \dots, N), \quad (13)$$

which implies that

$$\sum_{g=1}^N a_{g, \nu_g} = N^{1/2} \mathbf{t}' \bar{\mathbf{Y}}^{\text{obs}}.$$

By Assumption 1

$$\sum_{i'=1}^N a_{gi'} = 0, \quad (g = 1, \dots, N).$$

Therefore, if

$$\sum_{j'=1}^{j-1} n_{j'} < g \leq \sum_{j'=1}^j n_{j'},$$

then by (7) we have

$$b_{gi} = N^{1/2} t_j Y_i(z_j) / n_j - N^{-1/2} \sum_{j'=1}^J t_{j'} Y_i(z_{j'}). \quad (14)$$

The application of Lemma 1 hinges on (8), to prove which we consider two cases:

First we discuss the case in which “perfect co-linearity” does not hold, i.e., there exists $j \neq j'$ such that $\sigma_{jj'} < \sqrt{\sigma_{jj} \sigma_{j'j'}}$. On the one hand, (14) and Cauchy-Schwartz inequality imply that

$$\begin{aligned} b_{gi}^2 &\leq 2N t_j^2 Y_i^2(z_j) / n_j^2 + \frac{2}{N} \|\mathbf{t}\|_2^2 \sum_{j'=1}^J Y_i^2(z_{j'}) \\ &= \frac{2}{N} t_j^2 / \hat{p}_j^2 \sum_{i'=1}^N Y_{i'}^2(z_j) \frac{Y_i^2(z_j)}{\sum_{i'=1}^N Y_{i'}^2(z_j)} + \frac{2}{N} \|\mathbf{t}\|_2^2 \sum_{j'=1}^J \left\{ \sum_{i'=1}^N Y_{i'}^2(z_{j'}) \right\} \frac{Y_i^2(z_{j'})}{\sum_{i'=1}^N Y_{i'}^2(z_{j'})}, \end{aligned}$$

and consequently by (12) and Assumption 3

$$\lim_{N \rightarrow \infty} \max_{1 \leq g, i \leq N} b_{gi}^2 \leq 2 \left(\max_{1 \leq j \leq N} t_j^2 \sigma_{jj} / p_j^2 + \|\mathbf{t}\|_2^2 \sum_{j'=1}^J \sigma_{j'j'} \right) \times 0 = 0. \quad (15)$$

On the one hand, (14) implies that

$$\frac{1}{N} \sum_{g=1}^N \sum_{i=1}^N b_{gi}^2 = \sum_{j=1}^J t_j^2 / \hat{p}_j \left\{ \frac{1}{N} \sum_{i=1}^N Y_i^2(\mathbf{z}_j) \right\} - \sum_{j=1}^J \sum_{j'=1}^J t_j t_{j'} \left\{ \sum_{i=1}^N \frac{1}{N} Y_i(\mathbf{z}_j) Y_i(\mathbf{z}_{j'}) \right\},$$

and consequently by Assumptions 2 and 3

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{g=1}^N \sum_{i=1}^N b_{gi}^2 = \sum_{j=1}^J t_j^2 \sigma_{jj} / p_j - \sum_{j=1}^J \sum_{j'=1}^J t_j t_{j'} \sigma_{jj'}. \quad (16)$$

We prove the right hand side of (16) is always positive. By Cauchy-Schwartz inequality

$$\sum_{j=1}^J t_j^2 \sigma_{jj} / p_j = \left(\sum_{j=1}^J t_j^2 \sigma_{jj} / p_j \right) \left(\sum_{j=1}^J p_j \right) \geq \sum_{j=1}^J \sum_{j'=1}^J |t_j| |t_{j'}| \sqrt{\sigma_{jj} \sigma_{j'j'}}. \quad (17)$$

Because $\mathbf{t} \neq \mathbf{0}_J$, the equality sign in (17) holds if and only if $\mathbf{t} = \lambda(p_1/\sqrt{\sigma_{11}}, \dots, p_J/\sqrt{\sigma_{JJ}})'$ for a non-zero constant λ . Moreover, because $\sigma_{jj'} \leq \sqrt{\sigma_{jj} \sigma_{j'j'}}$ for all j and j' ,

$$\sum_{j=1}^J \sum_{j'=1}^J |t_j| |t_{j'}| \sqrt{\sigma_{jj} \sigma_{j'j'}} \geq \sum_{j=1}^J \sum_{j'=1}^J t_j t_{j'} \sigma_{jj'}.$$

Additionally, the fact that there exists $j_1 \neq j_2$ such that $\sigma_{j_1, j_2} < \sqrt{\sigma_{j_1, j_1} \sigma_{j_2, j_2}}$ implies that if $\mathbf{t} = \lambda(p_1/\sqrt{\sigma_{11}}, \dots, p_J/\sqrt{\sigma_{JJ}})'$, then

$$\sum_{j=1}^J \sum_{j'=1}^J |t_j| |t_{j'}| \sqrt{\sigma_{jj} \sigma_{j'j'}} > \sum_{j=1}^J \sum_{j'=1}^J t_j t_{j'} \sigma_{jj'}.$$

Thus we have proved that the right hand side of (16) is positive for all $\mathbf{t} \in \mathbb{R}^J \setminus \{\mathbf{0}_J\}$. Combining this fact with (15), we have proved that (8) holds for all $\mathbf{t} \in \mathbb{R}^J \setminus \{\mathbf{0}_J\}$, and therefore (9) holds by Lemma 1.

Second, we discuss the case in which “perfect co-linearity” holds, i.e., $\sigma_{jj'} = \sqrt{\sigma_{jj} \sigma_{j'j'}}$ for all j

and j' . If $\mathbf{t} \neq \lambda(p_1/\sqrt{\sigma_{11}}, \dots, p_J/\sqrt{\sigma_{JJ}})'$, similarly as the argument for the first case

$$\lim_{N \rightarrow \infty} \max_{1 \leq g, i \leq N} b_{gi}^2 = 0, \quad \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{g=1}^N \sum_{i=1}^N b_{gi}^2 > 0.$$

Therefore (8) holds, and consequently (9) holds by Lemma 1. Otherwise, by the definition of $\boldsymbol{\Sigma}^{\text{obs}}$,

$$\mathbf{t}' \boldsymbol{\Sigma}^{\text{obs}} \mathbf{t} = \lambda^2 \sum_{j=1}^J \left\{ (1 - p_j) \sqrt{\sigma_{jj}} - \sum_{j' \neq j} \frac{p_{j'} \sigma_{j'j}}{\sqrt{\sigma_{j'j'}}} \right\} = \lambda^2 \sum_{j=1}^J \left\{ (1 - p_j) \sqrt{\sigma_{jj}} - \sum_{j' \neq j} p_{j'} \sqrt{\sigma_{jj}} \right\} = 0.$$

Therefore $\text{Var}(N^{1/2} \mathbf{t}' \bar{\mathbf{Y}}^{\text{obs}}) \rightarrow 0$ by (11), and (9) holds trivially.

In summary, we have proved that (9) holds for all $\mathbf{t} \in \mathbb{R}^J$, which completes the proof. \square

Lemma 3. When $N \rightarrow \infty$,

$$\hat{\boldsymbol{\beta}}_j \xrightarrow{\mathbb{P}} \boldsymbol{\zeta}_j \quad (j = 1, \dots, J).$$

Proof. First, similarly as (10), for fixed N and $k = 1, \dots, p$,

$$\text{Var}\{\bar{X}_k^{\text{obs}}(\mathbf{z}_j)\} = \frac{1 - \hat{p}_j}{\hat{p}_j} \frac{1}{N(N-1)} \sum_{i=1}^N X_{ik}^2, \quad (18)$$

and

$$\begin{aligned} \text{Var} \left\{ \frac{1}{n_j} \sum_{i=1}^N W_i(\mathbf{z}_j) X_{ik} Y_i(\mathbf{z}_j) \right\} &= \frac{1 - \hat{p}_j}{\hat{p}_j} \frac{1}{N(N-1)} \sum_{i=1}^N \left\{ X_{ik} Y_i(\mathbf{z}_j) - \frac{1}{N} \sum_{i=1}^N X_{ik} Y_i(\mathbf{z}_j) \right\}^2 \\ &\leq \frac{2(1 - \hat{p}_j)}{(N-1)\hat{p}_j} \left[\frac{1}{N} \sum_{i=1}^N X_{ik}^2 Y_i^2(\mathbf{z}_j) + \frac{1}{N^2} \left\{ \sum_{i=1}^N X_{ik} Y_i(\mathbf{z}_j) \right\}^2 \right] \\ &\leq \frac{1 - \hat{p}_j}{\hat{p}_j} \frac{4L}{N-1}. \end{aligned} \quad (19)$$

The last step holds because by Cauchy-Schwartz inequality and Assumption 4

$$\frac{1}{N} \sum_{i=1}^N X_{ik}^2 Y_i^2(\mathbf{z}_j) \leq \left(\frac{1}{N} \sum_{i=1}^N X_{ik}^4 \right)^{1/2} \left\{ \frac{1}{N} \sum_{i=1}^N Y_i^4(\mathbf{z}_j) \right\}^{1/2} \leq L$$

and

$$\left\{ \sum_{i=1}^N X_{ik} Y_i(\mathbf{z}_j) \right\}^2 \leq \left(\sum_{i=1}^N X_{ik}^2 \right) \left\{ \sum_{i=1}^N Y_i^2(\mathbf{z}_j) \right\} \leq N \left(\sum_{i=1}^N X_{ik}^4 \right)^{1/2} \left\{ \sum_{i=1}^N Y_i^4(\mathbf{z}_j) \right\}^{1/2} \leq N^2 L.$$

Second, by (2) and (3), we only need to prove that when $N \rightarrow \infty$,

$$\frac{1}{n_j} \sum_{i=1}^N W_i(\mathbf{z}_j) \mathbf{X}_i \{Y_i(\mathbf{z}_j) - \bar{Y}^{\text{obs}}(\mathbf{z}_j)\} \xrightarrow{\mathbb{P}} \boldsymbol{\lambda}_j. \quad (20)$$

By (10) and (18) we have $\text{Var}\{\bar{Y}^{\text{obs}}(\mathbf{z}_j)\} \rightarrow 0$, and $\text{Var}\{\bar{X}_k^{\text{obs}}(\mathbf{z}_j)\} \rightarrow 0$ for all k . Therefore by Chebyshev inequality $\bar{Y}^{\text{obs}}(\mathbf{z}_j) \xrightarrow{\mathbb{P}} 0$ and $\bar{X}^{\text{obs}}(\mathbf{z}_j) \xrightarrow{\mathbb{P}} 0$. By (19)

$$\text{Var} \left\{ \frac{1}{n_j} \sum_{i=1}^N W_i(\mathbf{z}_j) X_{ik} Y_i(\mathbf{z}_j) \right\} \rightarrow 0 \quad (k = 1, \dots, p),$$

therefore

$$\frac{1}{n_j} \sum_{i=1}^N W_i(\mathbf{z}_j) \mathbf{X}_i Y_i(\mathbf{z}_j) \xrightarrow{\mathbb{P}} \boldsymbol{\lambda}_j,$$

and consequently (20) holds. \square

Lemma 4. When N approaches infinity, in distribution $N^{1/2} \bar{\mathbf{Y}}^{\text{ca}} \rightarrow N(0, \boldsymbol{\Sigma}^{\text{ca}})$, where

$$\boldsymbol{\Sigma}^{\text{ca}} = \begin{bmatrix} \frac{1-p_1}{p_1} \tilde{\sigma}_{11} & -\tilde{\sigma}_{12} & \dots & -\tilde{\sigma}_{1J} \\ -\tilde{\sigma}_{21} & \frac{1-p_2}{p_2} \tilde{\sigma}_{22} & \dots & -\tilde{\sigma}_{2J} \\ \vdots & \vdots & \ddots & \dots \\ -\tilde{\sigma}_{J1} & \dots & \dots & \frac{1-p_J}{p_J} \tilde{\sigma}_{JJ} \end{bmatrix}.$$

Proof. For $j = 1, \dots, J$, let $\bar{R}^{\text{obs}}(\mathbf{z}_j) = n_j^{-1} \sum_{i=1}^N W_i(\mathbf{z}_j) R_i(\mathbf{z}_j)$. By (4)

$$N^{1/2} \bar{\mathbf{Y}}^{\text{ca}} \stackrel{(4)}{=} \underbrace{N^{1/2} \begin{bmatrix} \bar{R}^{\text{obs}}(\mathbf{z}_1) \\ \vdots \\ \bar{R}^{\text{obs}}(\mathbf{z}_J) \end{bmatrix}}_{\Delta_1} - \underbrace{N^{1/2} \begin{bmatrix} \bar{\mathbf{X}}^{\text{obs}}(\mathbf{z}_1)'(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\zeta}_1) \\ \vdots \\ \bar{\mathbf{X}}^{\text{obs}}(\mathbf{z}_J)'(\hat{\boldsymbol{\beta}}_J - \boldsymbol{\zeta}_J) \end{bmatrix}}_{\Delta_2}.$$

On the one hand, $R_i(\mathbf{z}_j)$'s satisfy Assumption 4, because by Cauchy-Schwartz inequality

$$R_i(\mathbf{z}_j)^4 \leq (p+1)^3 \left\{ Y_i(\mathbf{z}_j)^4 + \left(\max_{1 \leq j \leq J} \|\boldsymbol{\zeta}_j\|_\infty^4 \right) \sum_{k=1}^p X_{ik}^4 \right\}.$$

By substituting $Y_i(\mathbf{z}_j)$ with $R_i(\mathbf{z}_j)$ and applying Lemma 2, we have $\Delta_1 \xrightarrow{\mathbb{D}} N(0, \boldsymbol{\Sigma}^{\text{ca}})$. On the other hand, by Lemma 3 we have $N^{1/2} \bar{\mathbf{X}}^{\text{obs}}(\mathbf{z}_j) = O_P(1)$ and $\hat{\boldsymbol{\beta}}_j - \boldsymbol{\zeta}_j = o_P(1)$, which implies that $\Delta_2 \xrightarrow{\mathbb{P}} 0$, and by Slutsky Theorem $\Delta_1 - \Delta_2 \xrightarrow{\mathbb{D}} N(0, \boldsymbol{\Sigma}^{\text{ca}})$. \square

4.3. Main Results

With the help of Lemmas 1–4, we now state and prove the main results.

Theorem 1. The randomization-based and covariate-adjusted estimators are both asymptotically normal, i.e.,

$$N^{1/2} \{\hat{\tau}_{\text{rb}}(l) - \tau(l)\} \xrightarrow{\mathbb{D}} N\{0, \sigma_{\text{rb}}^2(l)\}, \quad N^{1/2} \{\hat{\tau}_{\text{ca}}(l) - \tau(l)\} \xrightarrow{\mathbb{D}} N\{0, \sigma_{\text{ca}}^2(l)\},$$

where

$$\sigma_{\text{rb}}^2(l) = \frac{1}{2^{2(K-1)}} \left(\sum_{j=1}^J \frac{1-p_j}{p_j} \sigma_{jj} - \sum_{j \neq j'} h_{jl} h_{j'l} \sigma_{jj'} \right) \quad (21)$$

and

$$\sigma_{\text{ca}}^2(l) = \frac{1}{2^{2(K-1)}} \left(\sum_{j=1}^J \frac{1-p_j}{p_j} \tilde{\sigma}_{jj} - \sum_{j \neq j'} h_{jl} h_{j'l} \tilde{\sigma}_{jj'} \right). \quad (22)$$

of Theorem 1. The asymptotically normality of $\hat{\tau}_{\text{rb}}(l)$ follows from the fact that it is a linear combination of $\bar{\mathbf{Y}}^{\text{obs}}$, which by Lemma 2 is asymptotically multivariate normal. Moreover, (21) holds by (1). We apply similar argument to $\hat{\tau}_{\text{ca}}(l)$, in which we use Lemma 4. \square

Corollary 1. Let

$$\boldsymbol{\xi}_{jj'} = \left(\frac{p_{j'}}{p_j} \right)^{1/2} h_{jl} \boldsymbol{\zeta}_j - \left(\frac{p_j}{p_{j'}} \right)^{1/2} h_{j'l} \boldsymbol{\zeta}_{j'} \quad (j, j' = 1, \dots, J). \quad (23)$$

The difference of the asymptotic precisions between the randomization-based estimator and the

covariate-adjusted estimator is

$$\text{Var}\{\hat{\tau}_{\text{rb}}(l)\} - \text{Var}\{\hat{\tau}_{\text{ca}}(l)\} = \frac{1}{2^{2K-1}N} \sum_{j=1}^J \sum_{j'=1}^J \xi'_{jj'} \Omega \xi_{jj'}. \quad (24)$$

Proof of Corollary 1. On the one hand, by (23)

$$\xi'_{jj'} \Omega \xi_{jj'} = \frac{p_{j'}}{p_j} \zeta'_j \Omega \zeta_j + \frac{p_j}{p_{j'}} \zeta'_{j'} \Omega \zeta_{j'} - h_{jl} h_{j'l} \zeta'_j \Omega \zeta_{j'} - h_{j'l} h_{jl} \zeta'_{j'} \Omega \zeta_j. \quad (25)$$

On the other hand, by (21) and (22)

$$\begin{aligned} 2^{2(K-1)} \{\sigma_{\text{rb}}^2(l) - \sigma_{\text{ca}}^2(l)\} &= \sum_{j=1}^J \frac{1-p_j}{p_j} (\sigma_{jj} - \tilde{\sigma}_{jj}) - \sum_{j \neq j'} h_{jl} h_{j'l} (\sigma_{jj'} - \tilde{\sigma}_{jj'}) \\ &= \sum_{j=1}^J \frac{1-p_j}{p_j} \lambda'_j \Omega^{-1} \lambda_j - \sum_{j \neq j'} h_{jl} h_{j'l} \lambda'_j \Omega^{-1} \lambda_{j'} \\ &= \sum_{j=1}^J \frac{1}{p_j} \zeta'_j \Omega \zeta_j \sum_{j'=1}^J p_{j'} - \sum_{j=1}^J \sum_{j'=1}^J h_{jl} h_{j'l} \zeta'_j \Omega \zeta_{j'} \\ &= \frac{1}{2} \sum_{j=1}^J \sum_{j'=1}^J \xi'_{jj'} \Omega \xi_{jj'}. \end{aligned}$$

The last equation holds by (25). □

Theorem 1 illustrates the asymptotic unbiasedness and consistency of the randomization-based estimator and the covariate-adjusted estimator, and Corollary 1 illustrates the asymptotic precision by performing covariate adjustment. In particular, covariate adjustment never hurts asymptotic precision, and by (24) the sufficient and necessary condition for the randomization-based estimator and the covariate-adjusted estimator to be asymptotically equally precise is

$$p_{j'} h_{jl} \zeta_j = p_j h_{j'l} \zeta_{j'} \quad (j, j' = 1, \dots, J).$$

5. CONCLUDING REMARKS

In this paper, we define the covariate-adjusted estimator for 2^K factorial designs, and derive the asymptotic precisions of the unadjusted and covariate-adjusted estimators. We confirm that both

the unadjusted and covariate-adjusted estimators are asymptotically unbiased and normal, and the latter is more precise than the former. Moreover, we quantify the precision gained by performing covariate adjustment.

Our work implies multiple future directions. First, we can generalize our current framework to other factorial designs such as 3^k factorial designs or fractional factorial designs. Second, it is necessary to investigate the finite-sample properties of the estimators. In particular, although the covariate-adjusted estimator is asymptotically unbiased, it is biased from a finite-sample perspective. Lin (2013) showed that for randomized treatment-control studies the finite-sample bias of the covariate-adjusted estimator is $O(N^{-1})$, and it would be helpful to generate this result to factorial designs. Moreover, Lu (2016) showed that for 2^K factorial designs we can adopt the amended Huber-White sandwich estimator HC2 (MacKinnon and White 1985) for estimating the sampling variance of the unadjusted estimator, and therefore it would be helpful to have parallel results for the covariate-adjusted estimator. Third, it is possible to incorporate Bayesian analysis into our current framework.

ACKNOWLEDGEMENT

The author thanks several colleagues at the Microsoft Analysis and Experimentation Team, especially Randy Henne and Alex Deng, for inspiring this project, and Professor Tirthankar Dasgupta at Harvard University and Professor Peng Ding at UC Berkeley for helpful suggestions. Thoughtful comments from the Co-Editor-in-Chief, an Association Editor and a reviewer have substantially improved the quality of this paper.

REFERENCES

- Bailey, R. A. (1981). A unified approach to design of experiments. *Journal of the Royal Statistical Society: Series A*, 144(2):214–223.
- Bailey, R. A. (1991). Strata for randomized experiments (with discussion). *Journal of the Royal Statistical Society: Series B*, 53(1):27–78.
- Cochran, W. G. (1977). *Sampling Techniques, 3rd Edition*. New York: W.W. Norton.

- Dasgupta, T., Pillai, N., and Rubin, D. B. (2015). Causal inference from 2^k factorial designs using the potential outcomes model. *Journal of the Royal Statistical Society: Series B*, 77(4):727–753.
- Deng, A., Xu, Y., Kohavi, R., and Walker, T. (2013). Improving the sensitivity of online controlled experiments by utilizing pre-experiment data. In *Proceedings of the 6th ACM international conference on web search data mining*, pages 123–132.
- Espinosa, V., Dasgupta, T., and Rubin, D. B. (2016). A Bayesian perspective on the analysis of unreplicated factorial experiments using potential outcomes. *Technometrics*, 58:62–73.
- Fisher, R. A. (1935). *The Design of Experiments*. Edinburgh: Oliver and Boyd.
- Freedman, D. A. (2008). On regression adjustments in experiments with several treatments. *The Annals of Applied Statistics*, 2(1):176–196.
- Hoeffding, W. (1951). A combinatorial central limit theorem. *The Annals of Mathematical Statistics*, 22(4):558–566.
- Kempthorne, O. (1952). *The Design and Analysis of Experiments*. New York: Wiley.
- Kempthorne, O. (1955). The randomization theory of experimental inference. *Journal of American Statistical Association*, 50(271):946–967.
- Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman’s critique. *The Annals of Applied Statistics*, 7(1):295–318.
- Lu, J. (2016). On randomization-based and regression-based inferences for 2^k factorial designs. *Statistics and Probability Letters*, 112:72–78.
- Lu, J., Ding, P., and Dasgupta, T. (2015). Construction of alternative hypotheses for randomization tests with ordinal outcomes. *Statistics and Probability Letters*, 107:348–355.
- MacKinnon, J. G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325.
- Miller, S. (2006). *Experimental Design and Statistics*. London: Routledge.

- Miratrix, L. W., Sekhon, S. S., and Yu, B. (2013). Adjusting treatment effect estimates by post-stratification in randomized experiments. *Journal of the Royal Statistical Society: Series B*, 75(2):369–396.
- Neyman, J. S. (1990[1923]). On the application of probability theory to agricultural experiments. essay on principles (with discussion). section 9 (translated). reprinted ed. *Statistical Science*, 5(4):465–472.
- Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psuchology*, 66(5):688–701.
- Rubin, D. B. (1980). Comment on “Randomized analysis of experimental data: the fisher randomization test” by D. Basu. *Journal of American Statistical Association*, 75(371):591–593.
- Rubin, D. B. (2008). For objective causal inference, design trumps analysis. *The Annals of Applied Statistics*, 2(3):808–840.
- Wilk, M. B. and Kempthorne, O. (1956). Some aspects of the analysis of factorial experiments in a completely randomized design. *The Annals of Mathematical Statistics*, 27(4):950–985.
- Wu, C. F. J. and Hamada, M. S. (2009). *Experiments: Planning, Analysis, and Optimization*. New York: Wiley.
- Yates, F. (1937). The design and analysis of factorial experiments. *Technical Communication*, 35. Imperial Bureau of Soil Science, London.